1. start with $\alpha = \bar{y}$, $f_1 = f_2 = \ldots = f_p = 0$
2. For $j \in 1:p$
   a) get residual for $y_i - (\alpha + \sum_{k \neq j} f_k(x_k))$
   b) use linear smoother to regress on residuals (say we get function s_j)
   c) $f_j = s_j - \bar{s}_j$
3. repeat loop until things stop changing

Spline:

$$S_\lambda = \arg\min_m \frac{1}{n} \sum_{i=1}^{n} (y_i - m(x_i))^2 + \lambda \int \left(\frac{d^2 m}{dx^2}(x)\right)^2 dx$$

- computationally faster to fit spline than kernel

Hypothesis Testing with MSE

- Suppose we want to test if $\mu(x)$ is a linear function

H_0: $\mu(x)$ is a linear function H_a: $\mu(x)$ is not linear

Test statistic: D = difference in MSE

- If linear model is right it will predict at least as well as with non-parametric model
- If linear model is wrong, eventually non-parametric model will beat linear model

Need sampling dist of D under null hypothesis
- Simulate linear model and calculate D over and over (bootstrap)

- Find p-value and now you can reject or retain H_0 (if its linear)

— Random Note:
you can use hypothesis testing and the bootstrap to test anything like (for example is the noise gaussian?)

Can also do this:

Assuming ___ model, take residuals and run a smoother on residuals by x. Then hypothesis test on whether the smoother model is flat 0 or not (since it should be at 0 if model is true)

Logistic Regression

- We care about weighted $MSE = \frac{1}{n}\sum_{i=1}^{n} w_i(y_i - m(x_i))^2$
- In binary case we initialize $w_i = \frac{1}{Var[Y|X=x_i]} = \frac{1}{p(Y=1|X=x_i) - p(Y=1|X=x_i)^2}$